

RAW SEQUENCE LISTING

DATE: 11/27/2001

PATENT APPLICATION: US/09/986,682

TIME: 11:28:10

Input Set : N:\Crf3\RULE60\09986682.txt

Output Set: N:\CRF3\11272001\I986682.raw

SEQUENCE LISTING

3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: Tsusaki et al.

7 (ii) TITLE OF INVENTION: POLYPEPTIDE HAVING _-FRUCTOFURANOSIDASE

8 ACTIVITY

10 (iii) NUMBER OF SEQUENCES: 6

12 (iv) CORRESPONDENCE ADDRESS:

13 (A) ADDRESSEE: BROWDY AND NEIMARK

14 (B) STREET: 419 Seventh Street, N.W., Suite 300

15 (C) CITY: Washington

16 (D) STATE: D.C.

17 (E) COUNTRY: USA

18 (F) ZIP: 20004

20 (v) COMPUTER READABLE FORM:

21 (A) MEDIUM TYPE: Floppy disk

22 (B) COMPUTER: IBM PC compatible

23 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

24 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

26 (vi) CURRENT APPLICATION DATA:

C--> 27 (A) APPLICATION NUMBER: US/09/986,682

C--> 28 (B) FILING DATE: 09-Nov-2001

29 (C) CLASSIFICATION:

31 (vii) PRIOR APPLICATION DATA:

32 (A) APPLICATION NUMBER: 08/870,827

33 (B) FILING DATE: 06-JUN-1997

35 (A) APPLICATION NUMBER: JP 170,630/1996

36 (B) FILING DATE: 10-JUN-1996

38 (viii) ATTORNEY/AGENT INFORMATION:

39 (A) NAME: BROWDY, Roger L.

40 (B) REGISTRATION NUMBER: 25,618

41 (C) REFERENCE/DOCKET NUMBER: TSUSAKI=2

43 (ix) TELECOMMUNICATION INFORMATION:

44 (A) TELEPHONE: 202-628-5197

45 (B) TELEFAX: 202-737-3528

47 (2) INFORMATION FOR SEQ ID NO: 1:

49 (i) SEQUENCE CHARACTERISTICS:

50 (A) LENGTH: 21 amino acids

51 (B) TYPE: amino acid

52 (D) TOPOLOGY: linear

C--> 54 (v) FRAGMENT TYPE: N-terminal fragment

56 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

58 Met Asn Ser Gly Asp Tyr Lys Glu Asp Tyr Gly Phe Ala His Ile Thr

59 1 5 10 15

60 Arg Ala Asp Met Leu

61 20

63 (2) INFORMATION FOR SEQ ID NO: 2:

65 (i) SEQUENCE CHARACTERISTICS:

ENTERED

RAW SEQUENCE LISTING

DATE: 11/27/2001

PATENT APPLICATION: US/09/986,682

TIME: 11:28:10

Input Set : N:\Crf3\RULE60\09986682.txt

Output Set: N:\CRF3\11272001\I986682.raw

```

66          (A) LENGTH: 12 amino acids
67          (B) TYPE: amino acid
68          (D) TOPOLOGY: linear
C--> 70      (v) FRAGMENT TYPE: internal fragment
72      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
74 Ser Val Phe Asp Gly Gly Asp Gly Thr Val Tyr Gln
75 1          5          10
77 (2) INFORMATION FOR SEQ ID NO: 3:
79      (i) SEQUENCE CHARACTERISTICS:
80          (A) LENGTH: 455 amino acids
81          (B) TYPE: amino acid
82          (D) TOPOLOGY: linear
84      (ii) MOLECULE TYPE: peptide
86      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
88 Met Asn Ser Gly Asp Tyr Lys Glu Asp Tyr Gly Phe Ala His Ile Thr
89 1          5          10          15
90 Arg Ala Asp Met Leu Lys Ile Pro Gly Gln Gln Asn Ser Pro Gln Phe
91          20          25          30
92 Lys Val Pro Gln Phe Asn Ala Ser Ala Ile Lys Asn Ile Asp Ser Ala
93          35          40          45
94 Lys Gly Tyr Asp Lys Ser Gly Asn Leu Ile Asp Leu Asp Val Trp Asp
95          50          55          60
96 Ser Trp Pro Leu Gln Asn Ala Asp Gly Thr Ala Ala Asn Tyr His Gly
97 65          70          75          80
98 Tyr His Ile Val Ser Ala Leu Ala Gly Asp Pro Lys Asn Ser Asp Asp
99          85          90          95
100 Thr Pro Leu His Leu Phe Tyr Gln Lys Val Gly Asp Thr Ser Ile Asp
101          100          105          110
102 Ser Trp Lys Asn Ala Gly Arg Val Phe Glu Asp Met Asp Lys Phe Val
103          115          120          125
104 Pro Asn Asp Pro Tyr Leu Lys Tyr Gln Thr Gln Glu Trp Ser Gly Ser
105          130          135          140
106 Ala Thr Leu Thr Lys Asp Gly Gln Val Arg Leu Phe Tyr Thr Asp Tyr
107 145          150          155          160
108 Ser Gly Asn Pro Glu Asp Gly Gly Thr Gly Ala Gly Asn Gln Ile Ile
109          165          170          175
110 Ser Thr Ala Gln Val Asn Leu Ser Gln Pro Asp Ala Ala Thr Leu Lys
111          180          185          190
112 Val Asp Gly Val Ser Asp His Lys Ser Val Phe Asp Gly Gly Asp Gly
113          195          200          205
114 Thr Val Tyr Gln Asn Ile Gln Gln Phe Ile Asp Glu Gly Lys Trp Ile
115          210          215          220
116 Ser Gly Asp Asn His Thr Leu Arg Asp Pro His Tyr Val Glu Asp Lys
117 225          230          235          240
118 Gly His Lys Tyr Leu Val Phe Glu Ala Asn Thr Gly Thr Thr Asp Gly
119          245          250          255
120 Tyr Gln Gly Asp Gln Ser Phe Asn Asn Lys Ala Tyr Tyr Gly Gly Ser
121          260          265          270
122 Asp Val Phe Phe Gln Asn Glu Lys Asn Lys Leu Leu Gln Ser Pro Lys

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/986,682

DATE: 11/27/2001

TIME: 11:28:10

Input Set : N:\Crif3\RULE60\09986682.txt

Output Set: N:\CRF3\11272001\I986682.raw

```

123          275          280          285
124 Lys Gln Ile Ala Ser Leu Ala Asn Gly Ala Leu Gly Ile Val Glu Leu
125          290          295          300
126 Ala Asp Asp Tyr Thr Val Lys Ser Val Met Lys Pro Leu Val Ala Ser
127 305          310          315          320
128 Asn Thr Val Ala Asp Glu Val Glu Arg Ala Asn Ile Phe Lys Met Asn
129          325          330          335
130 Asn Lys Trp Tyr Leu Phe Thr Asp Ser Arg Gly Ser Lys Met Thr Ser
131          340          345          350
132 Asp Gly Ile Asn Asp Lys Asp Val Tyr Met Leu Gly Pro Gly Gly Asp
133          355          360          365
134 Ser Leu Asn Gly Pro His Asn Pro Ile Asn Glu Thr Gly Leu Val Leu
135          370          375          380
137 Asn Met Asn Leu Asp Pro Ala Asp Leu Thr His Thr Tyr Ser His Cys
138 385          390          395          400
139 Gly Ile Pro His Pro Glu Gly Asn Asn Val Val Leu Thr Ser Tyr Met
140          405          410          415
141 Thr Asn Arg Gly Phe Tyr Pro Glu His His Ser His Leu Arg Asp Lys
142          420          425          430
143 Leu Gly Val Asn Ile Lys Gly Ser Asp Thr Ser Gly Gly Glu Asn Ser
144          435          440          445
145 Ser Gly Gln Gly Gln Phe Pro
146          450          455
148 (2) INFORMATION FOR SEQ ID NO: 4:
150 (i) SEQUENCE CHARACTERISTICS:
151 (A) LENGTH: 1365 base pairs
152 (B) TYPE: nucleic acid
153 (C) STRANDEDNESS: double
154 (D) TOPOLOGY: linear
156 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
158 ATG AAC AGC GGG GAC TAC AAG GAA GAC TAT GGT TTT GCC CAT ATT ACA 48
159 Met Asn Ser Gly Asp Tyr Lys Glu Asp Tyr Gly Phe Ala His Ile Thr
160 1          5          10          15
162 CGC GCT GAC ATG CTA AAA ATT CCA GGA CAA CAA AAC AGT CCT CAA TTT 96
163 Arg Ala Asp Met Leu Lys Ile Pro Gly Gln Gln Asn Ser Pro Gln Phe
164          20          25          30
166 AAA GTG CCT CAA TTC AAT GCA TCA GCA ATC AAA AAC ATT GAT TCG GCA 144
167 Lys Val Pro Gln Phe Asn Ala Ser Ala Ile Lys Asn Ile Asp Ser Ala
168          35          40          45
170 AAA GGG TAT GAT AAG TCA GGC AAC TTA ATA GAT TTA GAT GTA TGG GAT 192
171 Lys Gly Tyr Asp Lys Ser Gly Asn Leu Ile Asp Leu Asp Val Trp Asp
172          50          55          60
174 AGC TGG CCA CTG CAA AAC GCT GAT GGT ACT GCG GCA AAT TAT CAT GGA 240
175 Ser Trp Pro Leu Gln Asn Ala Asp Gly Thr Ala Ala Asn Tyr His Gly
176 65          70          75          80
178 TAT CAC ATC GTC TCC GCT TTA GCA GGT GAC CCA AAA AAC AGT GAT GAT 288
179 Tyr His Ile Val Ser Ala Leu Ala Gly Asp Pro Lys Asn Ser Asp Asp
W--> 180          85          90          95
182 ACT CCA CTT CAT TTA TTC TAT CAA AAA GTC GGT GAT ACA TCG ATT GAC 336

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/986,682

DATE: 11/27/2001

TIME: 11:28:10

Input Set : N:\Crf3\RULE60\09986682.txt

Output Set: N:\CRF3\11272001\I986682.raw

```

183 Thr Pro Leu His Leu Phe Tyr Gln Lys Val Gly Asp Thr Ser Ile Asp
W--> 184 100 105 110 115
186 AGC TGG AAA AAT GCT GGA AGA GTA TTT GAA GAT ATG GAT AAA TTT GTT 384
187 Ser Trp Lys Asn Ala Gly Arg Val Phe Glu Asp Met Asp Lys Phe Val
W--> 188 120 125 130
190 CCA AAT GAT CCG TAT CTT AAA TAT CAA ACA CAG GAG TGG TCA GGT TCT 432
191 Pro Asn Asp Pro Tyr Leu Lys Tyr Gln Thr Gln Glu Trp Ser Gly Ser
W--> 192 135 140 145
194 GCT ACT TTA ACC AAA GAT GGC CAA GTC CGT TTA TTC TAT ACA GAT TAC 480
195 Ala Thr Leu Thr Lys Asp Gly Gln Val Arg Leu Phe Tyr Thr Asp Tyr
W--> 196 150 155 160
198 TCA GGT AAT CCT GAA GAT GGT GGA ACC GGT GCT GGT AAC CAA ATC ATT 528
199 Ser Gly Asn Pro Glu Asp Gly Gly Thr Gly Ala Gly Asn Gln Ile Ile
W--> 200 165 170 175
204 TCA ACT GCT CAA GTA AAC TTA TCC CAG CCG GAT GCA GCT ACA CTT AAA 576
205 Ser Thr Ala Gln Val Asn Leu Ser Gln Pro Asp Ala Ala Thr Leu Lys
W--> 206 180 185 190 195
208 GTC GAT GGA GTA TCT GAT CAT AAA TCT GTC TTT GAT GGC GGA GAC GGT 624
209 Val Asp Gly Val Ser Asp His Lys Ser Val Phe Asp Gly Gly Asp Gly
W--> 210 200 205 210
212 ACA GTT TAT CAA AAT ATT CAG CAA TTT ATC GAT GAA GGC AAG TGG ATT 672
213 Thr Val Tyr Gln Asn Ile Gln Gln Phe Ile Asp Glu Gly Lys Trp Ile
W--> 214 215 220 225
216 TCA GGT GAT AAC CAT ACT TTA AGA GAC CCT CAC TAT GTT GAA GAT AAG 720
217 Ser Gly Asp Asn His Thr Leu Arg Asp Pro His Tyr Val Glu Asp Lys
W--> 218 230 235 240
220 GGC CAT AAA TAT CTT GTC TTT GAA GCG AAT ACT GGA ACA ACA GAT GGT 768
221 Gly His Lys Tyr Leu Val Phe Glu Ala Asn Thr Gly Thr Thr Asp Gly
W--> 222 245 250 255
224 TAT CAA GGC GAT CAG TCT TTC AAT AAT AAA GCT TAC TAT GGC GGA AGT 816
225 Tyr Gln Gly Asp Gln Ser Phe Asn Asn Lys Ala Tyr Tyr Gly Gly Ser
W--> 226 260 265 270 275
228 GAC GTC TTC TTC CAG AAT GAA AAA AAT AAA CTG CTT CAA AGT CCT AAA 864
229 Asp Val Phe Phe Gln Asn Glu Lys Asn Lys Leu Leu Gln Ser Pro Lys
W--> 230 280 285 290
232 AAA CAA ATT GCT TCT TTA GCG AAT GGT GCA TTA GGC ATT GTT GAA TTG 912
233 Lys Gln Ile Ala Ser Leu Ala Asn Gly Ala Leu Gly Ile Val Glu Leu
W--> 234 295 300 305
236 GCC GAT GAC TAT ACA GTG AAA AGT GTT ATG AAA CCA TTA GTC GCA TCA 960
237 Ala Asp Asp Tyr Thr Val Lys Ser Val Met Lys Pro Leu Val Ala Ser
W--> 238 310 315 320
240 AAC ACA GTA GCA GAT GAA GTC GAA CGC GCC AAT ATA TTT AAA ATG AAT 1008
241 Asn Thr Val Ala Asp Glu Val Glu Arg Ala Asn Ile Phe Lys Met Asn
W--> 242 325 330 335
244 AAT AAA TGG TAT CTA TTC ACG GAT TCA AGA GGA TCC AAA ATG ACG AGT 1056
245 Asn Lys Trp Tyr Leu Phe Thr Asp Ser Arg Gly Ser Lys Met Thr Ser
W--> 246 340 345 350 355
248 GAT GGA ATT AAC GAC AAA GAT GTT TAT ATG CTA GGG CCC GGA GGC GAC 1104
249 Asp Gly Ile Asn Asp Lys Asp Val Tyr Met Leu Gly Pro Gly Gly Asp

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/986,682

DATE: 11/27/2001

TIME: 11:28:10

Input Set : N:\Crif3\RULE60\09986682.txt

Output Set: N:\CRF3\11272001\I986682.raw

```

W--> 250          360          365          370
      252 TCC TTA AAT GGC CCA CAC AAC CCG ATA AAT GAA ACT GGA CTT GTA TTG 1152
      253 Ser Leu Asn Gly Pro His Asn Pro Ile Asn Glu Thr Gly Leu Val Leu
W--> 254          375          380          385
      256 AAC ATG AAT CTT GAC CCT GCT GAT CTC ACA CAC ACT TAC TCT CAT TGC 1200
      257 Asn Met Asn Leu Asp Pro Ala Asp Leu Thr His Thr Tyr Ser His Cys
W--> 258          390          395          400
      260 GGT ATC CCG CAC CCT GAA GGT AAT AAT GTG GTA CTC ACA AGT TAT ATG 1248
      261 Gly Ile Pro His Pro Glu Gly Asn Asn Val Val Leu Thr Ser Tyr Met
W--> 262          405          410          415
      264 ACG AAT AGA GGC TTC TAT CCA GAA CAT CAC TCT CAC CTG CGG GAC AAG 1296
      265 Thr Asn Arg Gly Phe Tyr Pro Glu His His Ser His Leu Arg Asp Lys
W--> 266 420          425          430          435
      268 CTT GGG GTT AAT ATT AAA GGG TCT GAC ACA TCT GGA GGA GAA AAT AGT 1344
      269 Leu Gly Val Asn Ile Lys Gly Ser Asp Thr Ser Gly Gly Glu Asn Ser
W--> 270          440          445          450
      272 TCC GGA CAA GGA CAA TTC CCA 1365
      273 Ser Gly Gln Gly Gln Phe Pro
W--> 274          455
      276 (2) INFORMATION FOR SEQ ID NO: 5:
      278 (i) SEQUENCE CHARACTERISTICS:
      279 (A) LENGTH: 2408 base pairs
      280 (B) TYPE: nucleic acid
      281 (C) STRANDEDNESS: double
      282 (D) TOPOLOGY: linear
      284 (ii) MOLECULE TYPE: genomic DNA
      286 (vi) ORIGINAL SOURCE:
      287 (A) ORGANISM: Bacillus sp.
C--> 288 (C) INDIVIDUAL ISOLATE: V230 (FERM BP-5054)
      290 (ix) FEATURE:
      295 (A) NAME/KEY: signal peptide
      292 (B) LOCATION: 1..360
      293 (C) IDENTIFICATION METHOD: E
      307 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
      309 CGGGGAAAAT ACTAGATTCC AATTGGCCAG ACTTCCCAGT TGGTGTAAGA GAAGAGTTCTG 60
      310 GACTGCCAAT GCAGCTGTGC GTAAGAAAAC AGCTTACTCA TGAGCAATTA CTAGAAGAAT 120
      311 TTCAAAAAGTC CTGGGATAAG GCCAAGTCCA CTTTGAAATA AACTTTTCAG CCTCTGTGTG 180
      312 GGGGCTTTTT TGTTTTTATT TATTTCAACT GCAAGTGGTC CATCCCCTAT ATCAATTTAA 240
      313 GACGAAATTC TAATCAATCC ATGCCATCCC CAATAAACTC GTCCTCCTCT ATACTTTTAA 300
      314 TTAATAAGAA ACTATCAAGA GCTTCTTAT CAAATTCATA CATATCCAAG GAGGGAGACG 360
      315 ATG AAC TTC AAA AGA TTG GCG AAA AAA GCA GCT GCC GTA ACC TTC AGG 408
      316 Met Asn Phe Lys Arg Leu Ala Lys Lys Ala Ala Val Thr Phe Arg
      317 -30 -25 -20
      319 ACT GCT ATA TTA GTA GGA GCG GAC GGA CCG CAT ATT TTT GCG CAG CAA 456
      320 Thr Ala Ile Leu Val Gly Ala Asp Gly Pro His Ile Phe Ala Gln Gln
      321 -15 -10 -5
      323 ATG AAC AGC GGG GAC TAC AAG GAA GAC TAT GGT TTT GCC CAT ATT ACA 504
      324 Met Asn Ser Gly Asp Tyr Lys Glu Asp Tyr Gly Phe Ala His Ile Thr
      325 1 5 10 15

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/986,682

DATE: 11/27/2001

TIME: 11:28:11

Input Set : N:\Crf3\RULE60\09986682.txt

Output Set: N:\CRF3\11272001\I986682.raw

L:27 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:28 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:54 M:220 C: Keyword misspelled or invalid format, [(v) FRAGMENT TYPE:]
L:70 M:220 C: Keyword misspelled or invalid format, [(v) FRAGMENT TYPE:]
L:180 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:184 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:188 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:192 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:196 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:200 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:206 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:210 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:214 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:218 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:222 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:226 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:230 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:234 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:238 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:242 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:246 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:250 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:254 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:258 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:262 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:266 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:270 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:274 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4
L:288 M:220 C: Keyword misspelled or invalid format, [(C) INDIVIDUAL ISOLATE:]